AD_____

Award Number:
W81XWH-10-1-0589

TITLE:
Global Characterization of Protein-Altering Mutations in Prostate Cancer

PRINCIPAL INVESTIGATOR:
Jay A. Shendure, M.D., Ph.D.

CONTRACTING ORGANIZATION:
University of Washington, Seattle, WA 98195

REPORT DATE:
August 2013

TYPE OF REPORT:
Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
August 2013	Final	15 July 2010-14 July 2013

4. TITLE AND SUBTITLE

Global Characterization of Protein-Altering Mutations in Prostate Cancer

5a. CONTRACT NUMBER
W81XWH-10-1-0589

5b. GRANT NUMBER
W81XWH-10-1-0589

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jay Shendure

E-Mail: shendure@uw.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Washington
4333 Brooklyn AVE NE
Box 359472
Seattle, WA 98195-9472

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The research supported by this award aims to use a new generation of technologies for DNA sequencing to comprehensively scan the genomes of a series of prostate cancers for small mutations that disrupt protein-coding sequences. Our specific aims are as follows: (1) To carry out the genome-wide identification of nonsynonymous mutations in limited number of prostate metastases using second-generation technologies for targeted capture and sequencing; (2) To evaluate the mutational histories of individual mutations within the progression of the cancer in which it was observed, and to assess the prevalence of candidate cancer genes observed here in prostate cancer. (3) To perform integrative analyses of somatic mutation with gene expression and copy number change data collected on the same samples.

15. SUBJECT TERMS
Prostate cancer, genome, exome, sequencing, mutation, DNA

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	21	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

The identification of recurrent, protein-altering genetic alterations is frequently the means by which a given gene is initially implicated in tumor biology. However, we currently lack anything approaching a comprehensive picture of the protein-altering mutations that are biologically relevant or potentially specific to prostate cancer. The research supported by this award aims to use a new generation of technologies for DNA sequencing (Shendure and Ji 2008) to comprehensively scan the genomes of a series of prostate cancers for small mutations that disrupt protein-coding sequences. Our specific aims are as follows: (1) To carry out the genome-wide identification of nonsynonymous mutations in a limited number of prostate metastases using second-generation technologies for targeted capture and sequencing; (2) To evaluate the mutational histories of individual mutations within the progression of the cancer in which it was observed, and to assess the prevalence of candidate cancer genes observed here in prostate cancer. (3) To perform integrative analyses of somatic mutation with gene expression and copy number change data collected on the same samples.

Body

This is a "synergy" project between the laboratories of Dr. Jay Shendure in the Department of Genome Sciences at the University of Washington (UW) and Dr. Peter Nelson in the Division of Human Biology at the Fred Hutchinson Cancer Research Center (FHCRC). Because these are separate awards to the two investigators, this progress report is specific to tasks from the statement of work (SOW) assigned to the Shendure Lab only (or to progress within the Shendure Lab for joint tasks). Only tasks containing a UW component are listed here. Of note, Tasks 3, 4, 5, and 6 were largely performed in Year 1, whereas tasks 10, 11, 12, 13, 15, 16, 21 and 22 were largely performed in Years 2 and 3.

Aim 1: Perform a comprehensive screen for protein coding alterations in prostate metastases.

Task 3. DNA isolation and shotgun library construction (Months 1-10) [UW]

We performed exome sequencing of 23 prostate cancers derived from 16 different lethal metastatic tumors and 3 high grade primary carcinomas using solution-based hybrid capture (Nimblegen) followed by massively parallel sequencing (Illumina). Tumors were propagated in mice as xenografts. Genomic DNA was isolated from frozen tissue blocks using the QIAGEN DNeasy Blood and Tissue kit. Shotgun libraries were constructed by shearing gDNA, ligating sequencing adaptors, and performing PCR amplification.

Task 4. Array-based enrichment of coding sequences (Months 7-13) [UW]

The Nimblegen EZ SeqCap kit (Roche) was used as recently described (O'Roak, Deriziotis et al. 2011) in order to capture subsequences of the genome corresponding to coding regions, *i.e.* the "exome". Shotgun libraries were hybridized to either the EZ SeqCap V1 or V2 solution-based probes, and amplified. V1 probes (used in eight samples) targeted 26.6 Mb corresponding to the CCDS definitions of exons, while V2 probes (used in 15 samples) targeted 36.6 Mb corresponding to the RefSeq gene database.

Task 5. Massively parallel sequencing of tumor and control exomes (Months 10-16) [UW]

Post-enrichment libraries for these 23 prostate cancers were sequenced on either the Illumina GAIIx or HiSeq platforms (**Table 1**).

Table 1: Methods used to capture and sequence prostate cancer exomes. We used two versions of Nimblegen EZ SeqCap capture probes in this study. Eight samples were captured using V1 probes (targeting the 26.6 Mb Consensus Coding Sequence Database (CCDS), while the remainder of samples were captured using V2 probes (targeting the 36.6 RefSeq database). Four samples were indexed with barcodes prior to capture and sequencing. V1, Nimblegen V1 solution capture probes targeting CCDS coordinates; V2, Nimblegen V2 solution capture probes targeting RefSeq coordinates ; PE-76, paired-end sequencing using 76 bp reads; PE-100 paired-end sequencing using 100 bp reads.

Sample ID	Capture Method	Indexing	Sequencer	Run-type
LuCaP 23.1	V2	no	HiSeq	PE-100
LuCaP 23.12	V1	no	Illumina GAIIx	PE-76
LuCaP 23.1AI	V1	no	Illumina GAIIx	PE-76
LuCaP 35	V1	no	Illumina GAIIx	PE-76
LuCaP 35V	V1	no	Illumina GAIIx	PE-76
LuCaP 49	V1	no	HiSeq	PE-100
LuCaP 58	V2	no	HiSeq	PE-100
LuCaP 70	V2	no	HiSeq	PE-100
LuCaP 73	V2	yes	HiSeq	PE-100
LuCaP 77	V2	yes	HiSeq	PE-100
LuCaP 78	V2	no	HiSeq	PE-100
LuCaP 81	V2	no	HiSeq	PE-100
LuCaP 86.2	V1	no	HiSeq	PE-100
LuCaP 92	V2	no	HiSeq	PE-100
LuCaP 93	V2	no	HiSeq	PE-100
LuCaP 96	V1	no	Illumina GAIIx	PE-76
LuCaP 96AI	V1	no	Illumina GAIIx	PE-76
LuCaP 105	V2	no	HiSeq	PE-100
LuCaP 115	V2	no	HiSeq	PE-100
LuCaP 136	V2	no	HiSeq	PE-100
LuCaP 141	V2	no	HiSeq	PE-100
LuCaP 145.2	V2	yes	HiSeq	PE-100
LuCaP 147	V2	no	HiSeq	PE-100

Task 6. Read mapping, variant calling, and mutation annotation (Months 11-17) [UW]

We dealt with the possibility of mouse genomic DNA contamination by mapping sequence reads to both the human (UCSC hg18) and mouse (mm9) genome sequences using BWA (Li and Durbin 2009). Reads that mapped to the mouse genome were excluded from further analysis. See **Figures 1**, **2**, and **3**

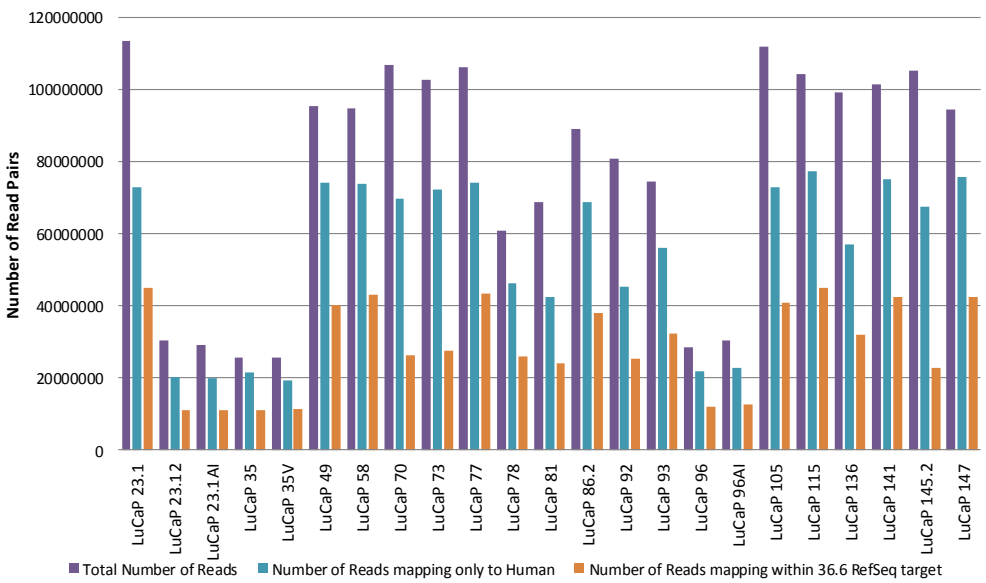

Figure 1: Summary of mapping statistics across 23 PCa xenograft exomes. LuCaP samples 96, 96AI, 23.12, 23.1AI, 35 and 35V were sequenced using the Illumina GAIIx, which accounts for the smaller number of reads obtained for these samples

for mapping statistics and calculations of mapping complexity. Sequence variant calls were performed by *samtools* (Li, Handsaker et al. 2009) after removing potential PCR duplicates, and were filtered to consider only positions with more than 8x coverage and a Phred-like consensus quality of 30 (Ng, Turner et al. 2009). To eliminate common germline polymorphisms from consideration, variants that had the same position as variants present in pilot data from the 1000 Genomes Project or in ~2,000 exomes corresponding to normal (non-tumor, non-xenografted) tissues sequenced at the University of Washington were removed from consideration. Genotypes were annotated using the SeattleSeq server (http://gvs.gs.washington.edu/SeattleSeqAnnotation/) and only nonsynonymous variants (missense/nonsense/splice-site mutations) were considered in identifying genes with recurrent mutations. The subset of genes that were recurrently mutated was then validated manually using IGV, the Integrated Genomics Viewer, to identify and remove false positive calls due to the presence of an insertion/deletion or incorrectly mapping read (12).

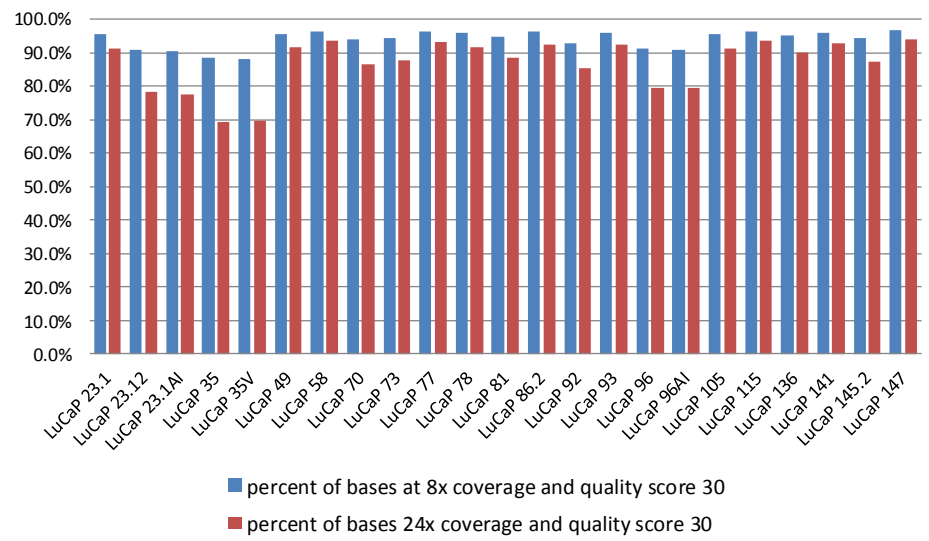

Figure 2: Fraction of bases in the V1 target definition that were covered to sufficient depth to enable basecalling.

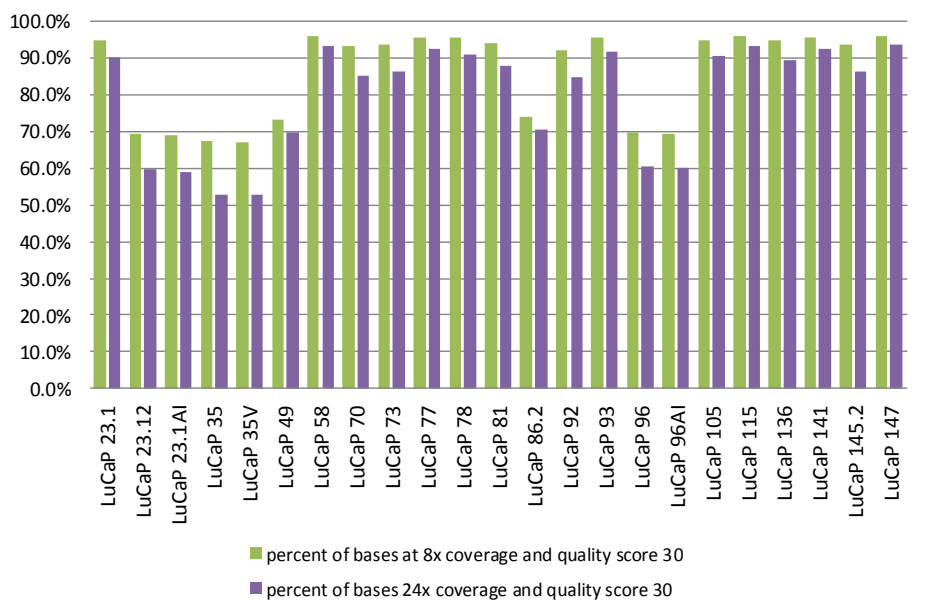

Figure 3: Fraction of bases in the V2 target definition that were covered to sufficient depth to enable basecalling. Samples LuCaP 96, 96V, 23.12, 23.1AI, 35, 35V, 49 and 86.2 were selected for a smaller (V1) target, which accounts for their relatively lower coverage of these regions.

SKP2	TP53	NRCAM	ITGA7
PTEN	EPHB2	BRCA2	PDZRN3
LRRK2	RAB32	ZNF473	GLI1
SPTA1	SF3A1	DKK1	DLK2
BRAF	ZFHX3	CHEK2	PCDH11X
TFG	TBX20	KLF6	PLXNB1
SPOP	AR	TAF1L	FOXA1
BDH1	MGAT4B	NMI	SDF4

Table 2: Genes attempted to sequence using Molecular Inversion Probe (MIP). A total of 32 genes were used in our first attempt of MIP sequencing. These genes were selected based on the results of a previous exome sequencing study as well as a literature review of mutations in prostate cancer.

Individual #	Sample #
Normal	55
Primary	17
Liver	19
Lymph node	67
Bone	13
Lung	12
Retroperitoneal	3
Spleen	2
Adrenal	2
Appendix	1
Peritoneum	3
Kidney	1
Skin	1

Table 3: Samples for MIP-based targeted resequencing

Aim 2: Evaluate mutational histories and prevalence screen of candidate cancer genes.

Task 10. Application to evaluate mutation histories (Months 20-24) [UW]
Task 11. Application to prevalence screen of candidate cancer genes (Months 18-30) [UW]

Our progress on Tasks 10 & 11 is discussed here jointly. For Aim 2, we identified a set of 32 candidate cancer genes on which to pursue further analysis both in terms of mutational history and prevalence through targeted resequencing (**Table 2**). The choice of what 33 genes to focus on was primarily based on the results of Aim 1 of this project (described above and also see **Table 6** below), although the list was supplemented with additional genes based on the literature in this area.

We designed molecular inversion probes (MIPs) corresponding to these genes and cost-effectively obtained these via massively parallel synthesis on and release from a DNA microarray (CustomArray). For each sample, the MIPs were added to 50 ng of DNA, followed by incubation with ligase, polymerase and nucleotides for 48 hours, resulting in targeted regions being "captured" within single-stranded circular DNA. After exonuclease removal of non-circularized DNA, captured products were amplified using PCR with barcoded primers containing adaptor sequences. The amplified products were pooled and sequenced on the HiSeq Illumina platform.

To date, we have performed MIP-based targeted resequencing of candidate cancer genes in 196 samples from 55 patients (**Table 3**). These include 55 normal DNAs (for comparison), 17 primary prostate cancers, and 124 metastatic prostate cancers from diverse sites.

With sequencing costs rapidly dropping, we also sought to pursue exome sequencing of additional prostate tumors as a more comprehensive approach for assessing mutation history and prevalence in established candidate cancer genes, with the added possibility of identifying additional candidate cancer genes. In close collaboration with the Nelson Lab, we contributed to the sequencing in Year 3 of an additional 180 exomes, corresponding to 13 primary tumors and 115 metastases derived from 52 patients (**Table 4**). For these more recent exomes, we used a

modified Nimblegen EZ SeqCap kit (Roche) protocol to capture subsequences of the genome that correspond to coding sequences. Shotgun libraries were first constructed with molecular barcodes attached to the adaptor sequences. These libraries were hybridized in pairs (two samples per capture reaction) to either the EZ SeqCap V2 or V3 solution-based probes, and amplified. V2 probes targeted 36.6 Mb corresponding to the RefSeq gene database while V3 probes targeted 64 Mb including Refseq gene database. Data collection was recently completed, and our analyses of the resulting data are described with Task 12.

Individual #	Sample #
Normal	50
Primary	14
Liver	19
Lymph node	61
Bone	15
Lung	11
Peritoneum	3
Spleen	1
Adrenal	2
Kidney	1
Appendix	1
Scrotum	1
Skin	1

Table 4: Summary of Samples for Additional Exome Resequencing

Task 12. Read mapping, variant calling, and mutation annotation (Months 21-31) [UW]

Sequencing data from MIP-based resequencing were aligned to the human hg19 reference genome using bwa and variant calls were made using SAMtools. The MIPs provided good coverage for 17 of the 33 target genes (**Table 5**; defined as >=100x coverage of >=65% of the targeted coding region). The remaining genes will likely require re-design and re-synthesis of MIPs, followed by another round of targeted capture and sequencing on the same samples.

	# of coding bases	% of coding bases with >8x coverage	% of coding regions with greater than 50x coverage	% of coding regions with greater than 100x coverage
SKP2	1447	98%	97%	96%
PTEN	1212	96%	96%	96%
LRRK2	7584	95%	94%	94%
SPTA1	7260	95%	93%	90%
BRAF	2301	93%	91%	90%
TFG	1203	89%	89%	89%
SPOP	1125	98%	96%	89%
NRCAM	3941	88%	88%	88%
BRCA2	10257	93%	90%	86%
ZNF473	2616	90%	88%	85%
DKK1	801	85%	85%	85%
CHEK2	1761	94%	91%	84%
KLF6	852	86%	86%	83%
TAF1L	5481	89%	86%	83%
NMI	924	98%	90%	82%
BDH1	1032	92%	79%	71%
TP53	1182	76%	69%	69%

Table 5: Coverage information across genes of interest. Of 32 genes initially screened for mutations, we obtained sufficient coverage (>100x coverage in >65% of coding bases) for 17 genes. We are currently redesigning probes to capture the remaining portions of coding regions.

Positions were considered to be called at high quality if they had at least 8x coverage and a Phred-scaled quality score of 30. A list of mutations identified to date in these genes via MIP based targeted resequencing is provided further below in the context of **Task 15 & 16**.

Read-mapping, variant calling, and mutation annotation of the additional 180 exomes sequenced in Year 3 was performed as follows. BWA was used to align reads to the 1000 genomes reference (hs37d5) and GATK was used for local realignment (**Figure 4**). SAMtools was used to remove duplicates and sort and index read files. Mutations were called using Mutect with standard parameters. To deal with potential barcode crosstalk that may have occurred due to pairing of samples before exome capture, we removed all variants in the paired sample that were also present in the original sample. This process left a total of 23,270 mutations and 13,381 coding mutations across all individuals. Mutational analysis of these data is provided below in the context of **Tasks 15 & 16.**

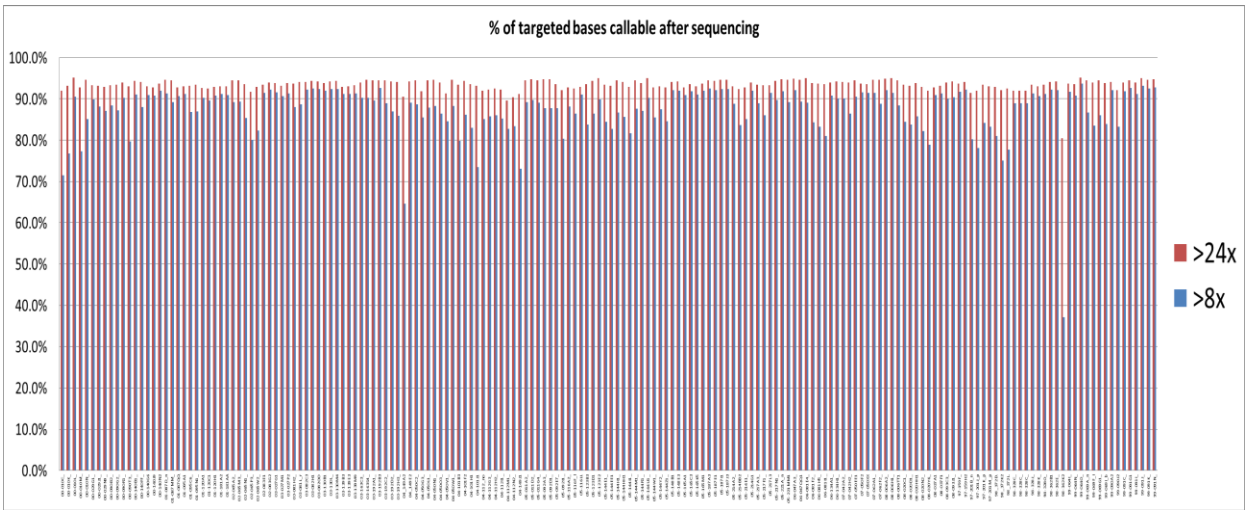

Figure 4: Fraction of bases in the V2 target definition that were covered to sufficient depth to enable basecalling.

Task 13. Verification/confirmation of sequence variants (Months 20-26) [UW]

Validation of mutations identified by targeted resequencing of candidate cancer genes is ongoing. We have prioritized the validation of mutations with potential clinical significance. For example, we validated a K601E mutation in BRAF in all metastases of a single patient by Sanger sequencing (**Figure 5**). This residue is immediately adjacent to the residue that is most commonly mutated in BRAF (V600E). As there are therapeutic agents specific to tumors with BRAF mutations in the context of other cancers (e.g. melanoma), this finding has potential

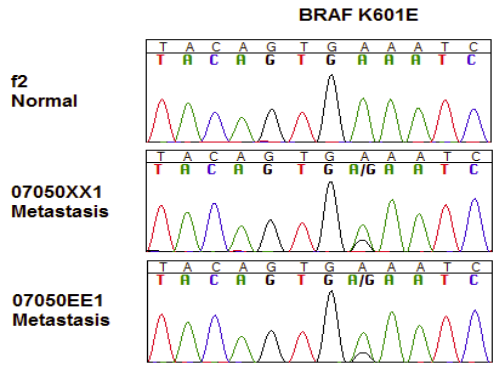

Figure 5. Sanger Traces of the BRAF exon containing the K601E mutation in two representative bone metastases. BRAF mutations were subsequently confirmed in 8/8 bone metastases from this individual.

clinical implications in that it may indicate that a small but appreciable fraction of prostate cancer patients may have BRAF mutations and will potentially be responsive to these same agents.

A second set of mutations for which validations are in progress in a patient in whom different AR mutations were observed. Specifically, subsets of metastases in patient #47 had either a AR point mutation or AR amplification in a mutually exclusive fashion (**Table 5**). To our knowledge, such intra-patient heterogeneity with respect to AR mutation has not been previously documented.

Patient ID	Metastatic site	AR point mutation	AR amplification
47	Normal liver	--	--
47	bladder	H874Y	---
47	prostate	H874Y	**---**
47	R pericaval LN	---	High Copy
47	R pericaval LN #2	---	High Copy
47	R periaortic LN #3	---	High Copy

Table 5. Intra-patient heterogeneity of AR mutations. Point mutation and Copy Number status of AR was assessed using Molecular Inversion Probe technology and is current being validated with other methods. One patient (47) displayed two different forms of mutation in AR following treatment with ADT and estrogen treatment.

Aim 3: Integrate analyses of molecular alterations in metastatic and primary prostate cancer.

Task 15. Analysis of mutational patterns in metastatic prostate cancer (Months 11-17) [UW]

Task 16. Analysis of mutational patterns in primary prostate cancer (Months 21-31) [UW]

To date, our analyses of mutational patterns have produced two key results. First, we have found that a subset of prostate cancers exhibit a "hypermutator" phenotype with respect to point substitution mutations. Second, we have identified a set of candidate genes that may be recurrently mutated in prostate cancer. These analyses are discussed separately below.

Prostate cancers with "hypermutated" genomes. We observed that the exomes of three prostate cancers, LuCaP 58, LuCaP 73 and LuCaP 147 possessed a strikingly high number of novel, non-synonymous single nucleotide variants, nearly tenfold more than other tumors (p=0.0097) (**Figure 6**). There were no distinctive features to suggest why these tumors should have more variants. Each tumor originated as a high grade Gleason 9 cancer, all were from individuals of Caucasian ancestry, one represented a primary neoplasm, one a lymph node metastasis, and one a metastasis to the liver. We hypothesized that the large number of nov-SNVs observed in three prostate cancers may be due to a 'mutator phenotype' that developed during the initial stages of tumorigenesis, as a consequence of therapeutic pressures and subsequent clonal selection, or evolved while passaged in the mouse hosts. To determine if these results reflect truly elevated numbers of somatic mutations within human tumors and not as a result of passage within mice, we sequenced paired normal and directly resected, non-xenografted, tumor samples corresponding to one hypermutated xenograft (LuCaP 147), and two non-hypermutated xenograft lines (LuCaP 92 and LuCaP 145.2). Of 2,368 novSNVs in LuCaP147

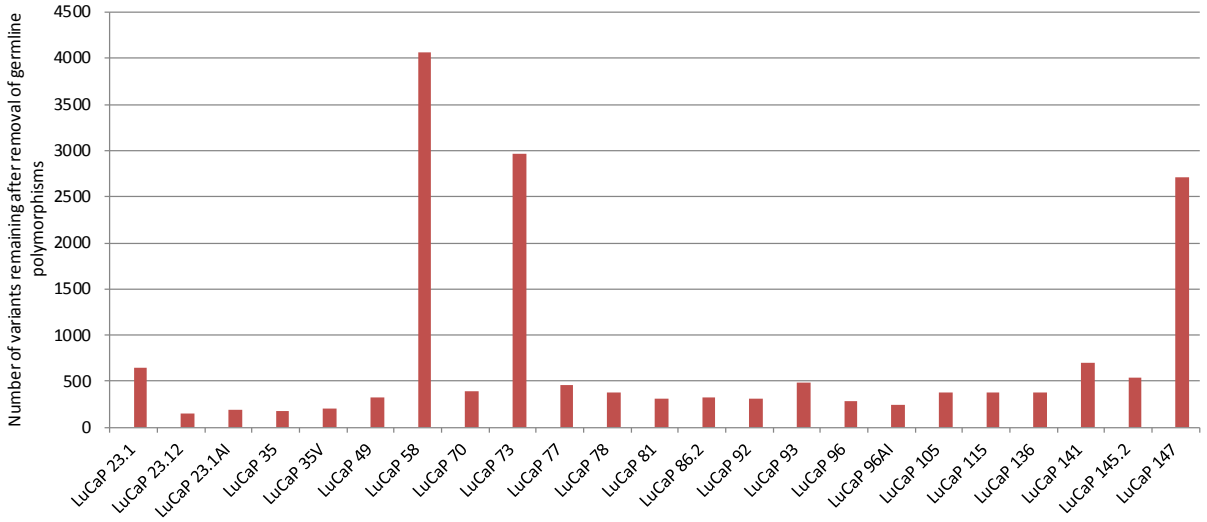

Fig. 4. A subset of xenografts exhibit a high number of mutations. After filtering to remove common germline polymorphisms, three xenografts (LuCaP 73, LuCaP 147 and LuCaP 58) exhibit a "hypermutated" phenotype, with several thousand novel SNVs each. This contrasts with the other 20 xenografts, which have 362 +/- 147 coding alterations remaining after filtering.

able to be called across all three samples (xenograft, derivative tumor and normal tissue) 1,402 were somatic and present with metastasis tissue. In contrast, the other two non-xenografted tumors (corresponding to LuCaP 92 and LuCaP 145.2) had between 31 and 58 somatic mutations. Furthermore, because we sequenced a neighboring metastasis, rather than the exact metastasis from which LuCaP147 was derived, the result indicates that at least these ~1,400 somatic mutations were shared between these metastases. The vast majority of the ~1,300 novSNVs observed in the LuCaP147 xenograft but not the metastasis likely occurred during passaging within mice, or were specific to the metastasis from which LuCaP147 was derived.

As discussed above, we have more recently contributed to the sequencing in Year 3 of an additional 180 exomes, corresponding to 13 primary tumors and 115 metastases derived

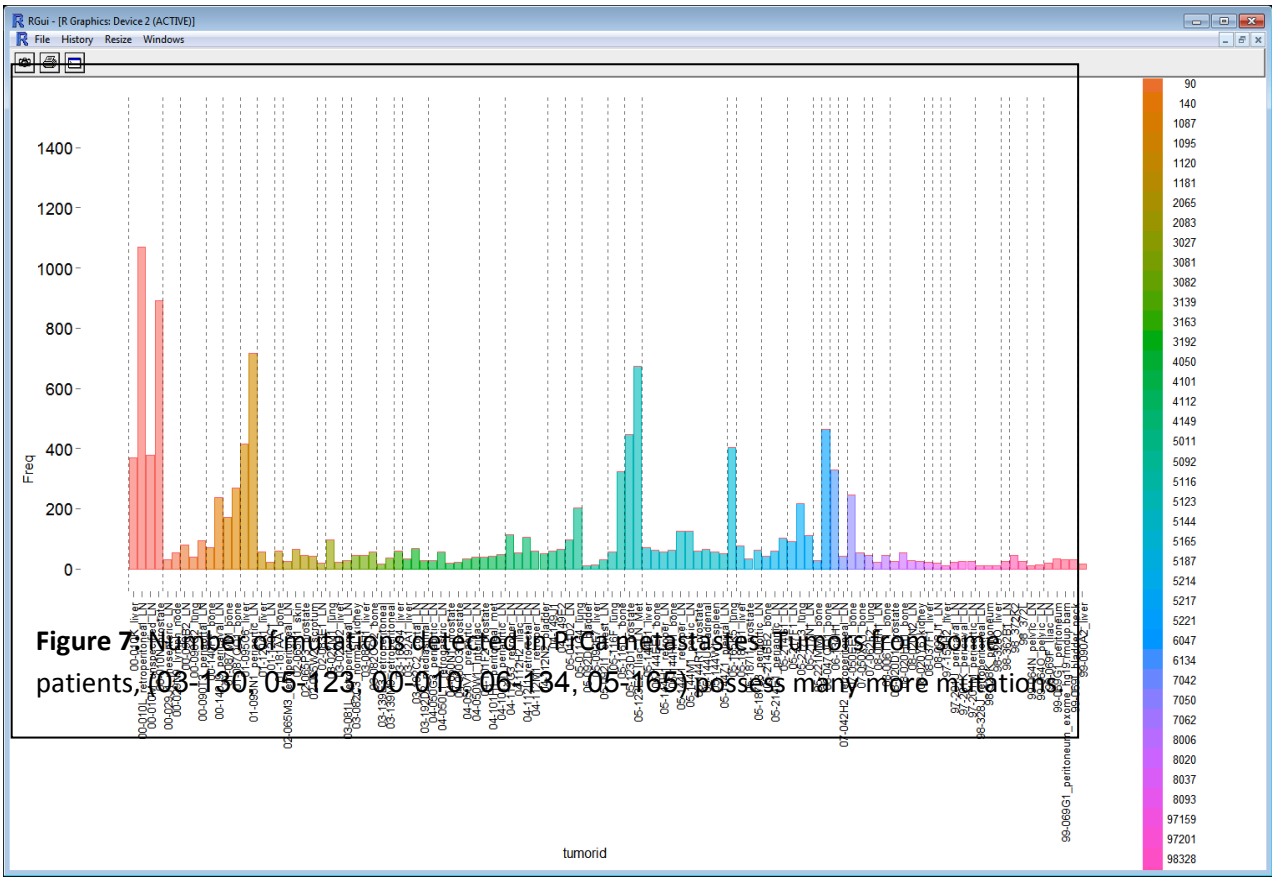

Figure 7: Number of mutations detected in PrCa metastases. Tumors from some patients, (03-130, 05-123, 00-010, 06-134, 05-165) possess many more mutations

from 52 patients. Of these, 6 (03-130, 05-123, 00-010, 06-134, 05-165, and 01-095) had tumors that were hypermutated (**Figure 7**). This brings the total number of hypermutated tumors up to 8 (one of these tumors was previously observed to be hypermutated in our earlier study). We have begun to collaborate with Colin Pritchard (Lab Medicine, UW) to investigate the possible role of mismatch repair defects in the development of hypermutated tumors.

Mutations were integrated with data from three additional exome studies, resulting in a meta-analysis of more than 450 patients with PrCa and >100 patients with CRPC. Because our experimental design involved sequencing multiple tumors from the same individual, we could begin to look at how different tumors from the same patient share mutations in clinically associated genes. To conduct this analysis, we first curated a list of 20 genes that had been previously identified as significantly mutated or associated with prostate cancer. By combining point mutation (from exomes) and copy number information (from Agilent array run at the Nelson lab) we produced the first integrated picture of intrapatient mutational heterogeneity across many individuals with prostate cancer (**Figure 8**). Most tumors shared mutations in known driver genes, however there were instances of disagreement. Notable cases of heterogeneity included three cases of heterogeneity of AR amplification, with one case of AR inactivating mutation present in one tumor but not the other. We also identified one case of PTEN mutation in 98-362 that was not present in another metastasis.

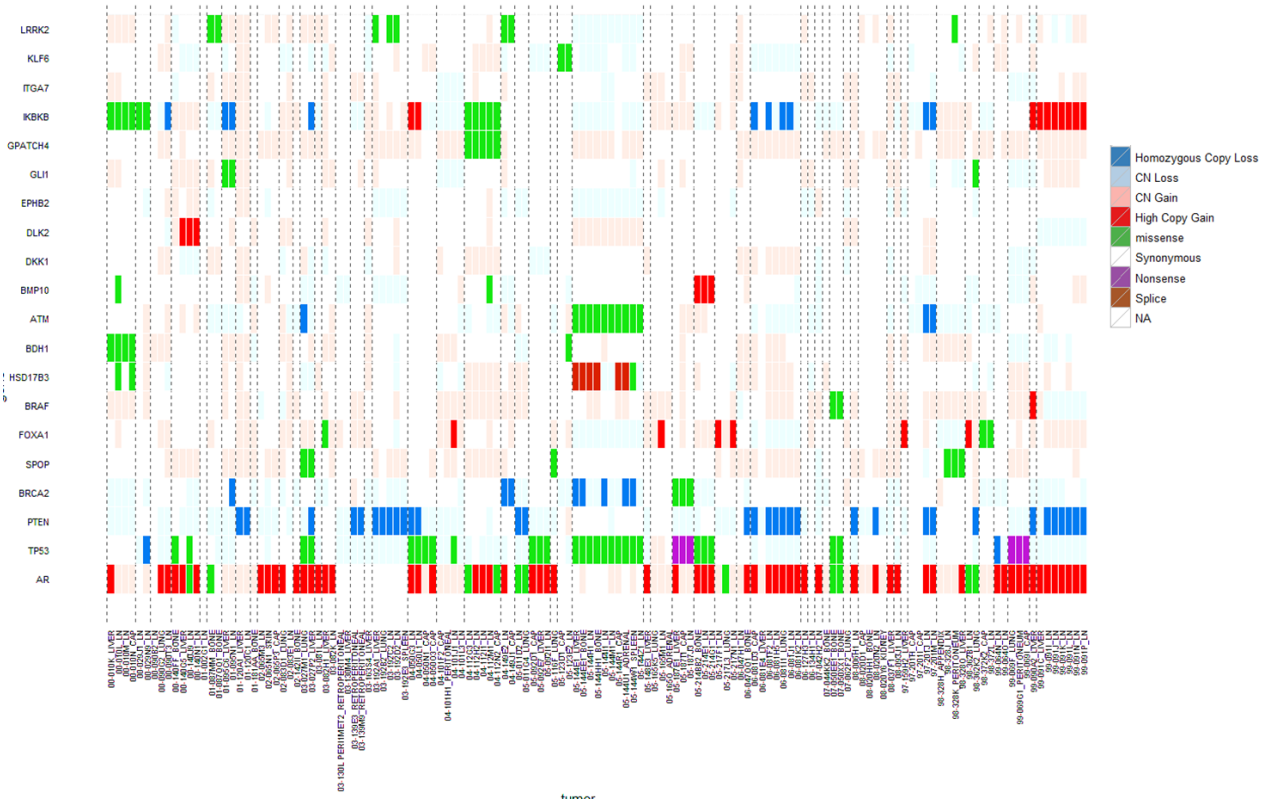

Figure 8: Extent of mutational heterogeneity in clinically relevant genes across 50 patients with 13 primary tumors. Shown are patients with primary and metastasis sequenced and mutations in common genes shown. Dashed lines separate tumors from different individuals. Genes were considered to be mutated if at least one sample was mutated. Sites are limited to those with mutations in COSMIC samples. Many tumors later develop therapy associated mutations.

Recurrent nonsynonymous genomic sequence alterations in prostate cancers. We examined the set of novel, nonsynonymous single nucleotide variants (nov-nsSNVs) to identify those genes that may be recurrently affected by protein-altering point mutations across different tumors. In order to reduce spurious findings due to inconsequential passenger mutations, we excluded the three "hypermutated" tumors from this analysis. We also manually examined read pileups for variants in genes with potential recurrence attributable to basecalling artifacts due to either insertions/deletions or poorly mapping reads. Across 16 tumors from unrelated individuals, 131 genes had nov-nsSNVs in two or more exomes, and 23 genes had nov-nsSNVs in three or more exomes. A subset of the novel variants are likely due to instances where very rare germline variants (i.e. not seen in several thousand other chromosomes) occur in the same gene, as we cannot distinguish these from somatic mutations. We therefore excluded from consideration the 1% of genes with the highest rate of very rare germline variants, i.e. singletons, based on an analysis of control exomes (as some genes are much more likely to contain very rare germline variants than other genes) (Bustamante, Fledel-Alon et al. 2005; Lohmueller, Indap et al. 2008). This reduced the number of candidates to 104 genes with nov-nsSNVs in two or more exomes, and 12 genes with nov-nsSNVs in three or more exomes (**Figure 9**). To further segregate candidate genes with the goal of identifying those with recurrent somatic mutations, we estimated the probability of recurrently observing germline nov-nsSNVs in each candidate gene by iterative sampling from 1,865 other exomes sequenced at the University of Washington. We excluded from consideration genes for which the probability of observing the genes recurrently mutated due to germline variation was greater than 0.001. This reduced the number of candidates to 20 genes with nov-nsSNVs in two or more exomes, and 10 genes with nov-nsSNVs in three or more exomes (**Table 6**). Notably, whereas we began with 4 genes with nov-nsSNVs in four or more exomes (MUC16, SYNE1, UBR4, and TP53), only one of these (TP53) remained in our final candidate list, where it is the most significant. These data and analysis provide a strong set of candidates for further investigation.

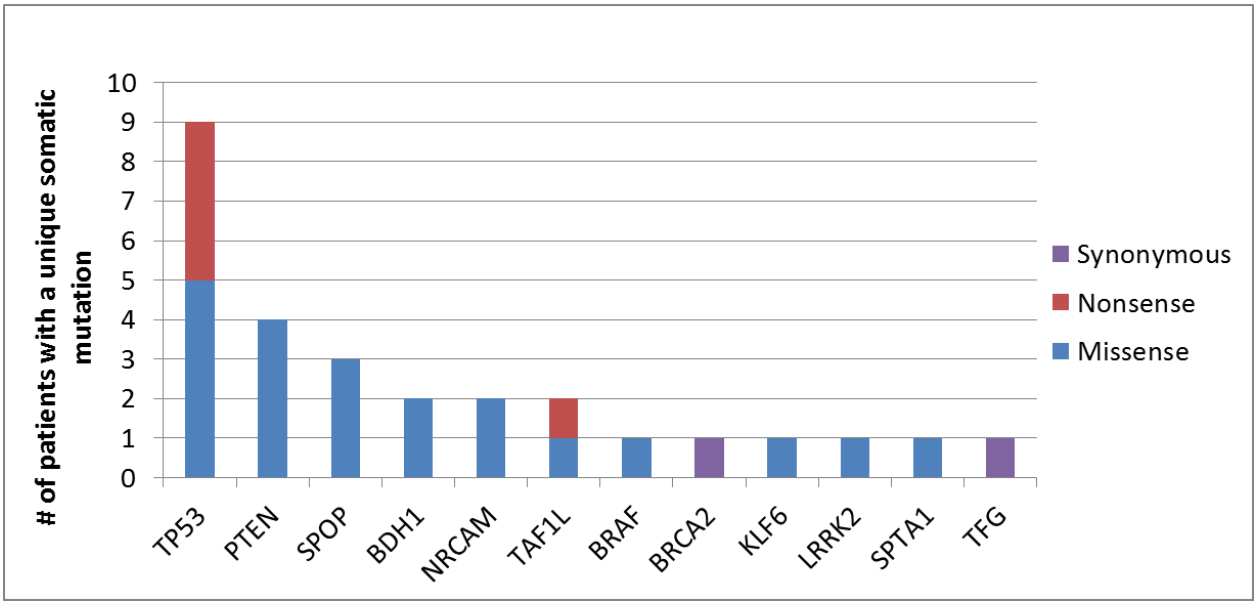

Figure 9: Summary of genes with somatic mutations in prostate cancer. Using a strict cut off of 100x coverage across samples, we identified somatic mutations in 12/17 genes investigated. TP53 shows the highest rate of mutation, followed by PTEN, BDH1 and SPOP.

# of samples	Gene ID	Gene Name	P-value of being germline	Individual mutations seen
5	TP53	tumor protein p53 (Li-Fraumeni syndrome)	< 0.00005	LuCaP73(ARG306GLN), LuCaP136(ARG280stop), LuCaP96AI(CYS238TYR), †LuCaP92(GLU198stop), LuCaP73(ARG175CYS), LuCaP70(TYR163HIS), LuCaP77(PRO278SER)
3	SDF4	stromal cell derived factor 4	< 0.00005	LuCaP108(ASP276ASN), LuCaP78(GLY76SER), LuCaP115(ALA9SER)
3	PDZRN3	PDZ domain containing RING finger 3	< 0.00005	LuCaP96AI(ARG727CYS), LuCaP108(GLY570SER), LuCaP73(ARG463CYS), LuCaP92(ILE331LEU)
3	DLK2	delta-like 2 homolog	0.00005	LuCaP70(ARG371HIS), *LuCaP145.2(SER361ARG), LuCaP23.1AI(HIS280GLN)
3	FSIP2	fibrous sheath interacting protein 2	0.00005	LuCaP81(LYS22ASN), †LuCaP92(THR698ILE), LuCaP136(GLN1526HIS)
3	NRCAM	neuronal cell adhesion molecule	0.00015	LuCaP115(MET1094ILE), LuCaP86.2(LYS645GLU), †LuCaP145.2(SER329CYS)
3	MGAT4B	mannosyl GAT4B	0.0002	LuCaP108(ALA504THR), LuCaP96AI(ARG168CYS), LuCaP136(VAL150MET)
3	PCDH11X	protocadherin 11 X-linked	0.0003	*LuCaP145.2(VAL38PHE), LuCaP58(MET867VAL), LuCaP108(VAL1007ILE), LuCaP49(THR1296ASN)
3	GLI1	glioma-associated oncogene homolog 1 (zinc finger protein)	0.0003	LuCaP86.2(ARG20TRP), LuCaP78(ARG81GLN), LuCaP23.1AI(PRO210THR)
3	KDM4B	Lysine-specific demethylase 4B	0.00035	LuCaP73(ALA265VAL), LuCaP108(ARG534TRP), LuCaP35V(ALA555VAL), LuCaP73(ALA827VAL), LuCaP86.2(SER1036CYS)
2	DKK1	dickkopf homolog 1	< 0.00005	†LuCaP92(GLU151GLN), LuCaP93(SER244TYR)
2	RAB32	RAB32, member RAS oncogene family	0.00005	LuCaP93(VAL66ILE), LuCaP141(SER109stop)
2	PLA2G16	phospholipase A2, group XVI	0.00015	LuCaP115(SER85LEU), LuCaP35V(PRO19HIS)
2	TFG	TRK-fused gene	0.00015	LuCaP96AI(ASN134HIS), LuCaP141(GLN318stop), LuCaP147(TYR319stop)
2	TBX20	T-box 20	0.0002	LuCaP77(ARG437HIS), LuCaP23.1AI(ALA52SER)
2	ZNF473	zinc finger protein 473	0.00025	LuCaP108(VAL465ILE), LuCaP115(GLY652ARG)
2	SF3A1	splicing factor 3a, subunit 1, 120kDa	0.0006	LuCaP70(PRO558LEU), LuCaP23.1AI(VAL479ILE)
2	NMI	N-myc (and STAT) interactor	0.00075	LuCaP141(ILE302ARG), LuCaP86.2(GLN101ARG)
2	IKZF4	IKAROS family zinc finger 4 (Eos)	0.0008	LuCaP93(ASP106ASN), LuCaP81(ASP498ASN)
2	BDH1	3-hydroxybutyrate dehydrogenase, type 1	0.00095	LuCaP73(VAL190ILE), LuCaP96AI(THR176MET), LuCaP147(VAL142ILE), LuCaP115(HIS74TYR), LuCaP147(ALA50VAL)

Table 6: Genes with recurrent novel, nonsynonymous alterations. P-values were estimated by randomly sampling from 1,865 other exomes sequenced at the University of Washington to estimate the probability of recurrently observing nov-nsSNVs in a given candidate gene. These are the 20 genes with the best p-values. *The nov-nsSNV in this gene was determined to be a rare germline mutation within this xenograft. †The nov-nsSNV in this gene was determined to be a somatic within this xenograft.

As discussed above, we have more recently contributed to the sequencing in Year 3 of an additional 180 exomes, corresponding to 13 primary tumors and 115 metastases derived from 52 patients. Mutations were integrated with data from three additional exome studies, resulting in a meta-analysis of more than 450 patients with PrCa and >100 patients with CRPC. In addition to looking at overall patterns of mutation, we also looked at genes with point mutations that are recurrent or appear to cluster. Our initial analyses have revealed known genes, i.e. TP53, AR, and SPOP, along with additional candidates that we are following up on. We have also looked at genes mutated more often in cases of castration resistant prostate cancer (CRPC) but not within primary prostate tumors (**Table 7**). We have thus far found multiple CRPC-associated genes including some that were previously reported (AR, ZFHX3, MLL2) and others that are new. Our hypothesis is that these genes are important for later stages of PrCa progression including metastasis and resistance to treatment.

Gene Code	Mut. Freq in our study of (50, CRPC)	Mut. Freq in Grasso et al (50, CRPC)	Mut. Freq in Bariberi et al (110, Primary)	Mut. Freq in Lindberg et al (50, Primary)	Mut. Frequency in TCGA (200, Primary)
AR	8.9%	8.1%	0.0%	0.0%	0.0%
Candidate A	11.1%	4.8%	0.0%	0.0%	1.7%
ZFHX3	8.9%	6.5%	0.9%	0.0%	0.6%
Candidate B	6.7%	4.8%	0.0%	0.0%	0.0%
Candidate C	6.7%	4.8%	0.0%	0.0%	2.3%
Candidate D	6.7%	4.8%	0.9%	0.0%	0.0%
MLL2	6.7%	4.8%	0.9%	0.0%	1.1%
Candidate E	4.4%	4.8%	0.0%	1.6%	0.0%

Table 7: Genes preferentially mutated in CRPC tumors. We compared the mutation frequency of genes mutated in studies of Castration Resistant Prostate Cancer (CRPC) with the mutation frequency of genes in studies of localized primary prostate cancer. This analysis resulted in a number of genes mutated in castration resistant prostate cancer but not in other tumors. This table summarizes the most CRPC-associated genes with their frequency across other studies.

Task 22. Completing project reports and manuscripts (Months 11-36) [UW & FHCRC]

A portion of the work described in this progress report was published in the Proceedings of the National Academy of Sciences (PNAS) in September 2011:

Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, Rieder MJ, Nickerson DA, Corey E, Lange PH, Morrissey C, Vessella RL, Nelson PS, Shendure J. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci U S A. 2011 Oct 11;108(41):17087-92. Epub 2011 Sep 26. PubMed PMID: 21949389; PubMed Central PMCID: PMC3193229.

In collaboration with the Nelson Lab, we are currently preparing two additional manuscripts. One is a case report on the observation of heterogeneity of AR mutations within a patient. The other is a manuscript focused on mutational history in individual patients as well as on the prevalence of mutations in candidate cancer genes, including both data from MIP based resequencing and the exome sequencing of 128 additional tumors.

Key Research Accomplishments

- We have performed high-quality whole exome sequencing of 23 prostate cancers derived from 16 different lethal metastatic tumors and 3 high grade primary carcinomas.

- We have found that a subset of prostate cancers that exhibit a clear "hypermutator" phenotype with respect to point mutations, with potential implications for resistance to cancer therapeutics. This finding has been subsequently verified by several groups.

- We have performed a prevalence screen for somatic mutations in 17 genes on prostate cancer samples from 55 patients, including 17 primary prostate cancers, and 124 metastatic prostate cancers from diverse sites.

- We have identified two interesting cases with potential therapeutic implications, including one case of a patient with BRAF mutations in all prostate cancer metastases, and a second case of a patient with heterogeneity with respect to the nature of their AR mutation driving therapeutic resistance in different metastases.

- We contributed to the sequencing in Year 3 of an additional 180 exomes corresponding to 13 primary tumors and 115 metastases derived from 52 patients. Analysis of these data with respect to both mutational history in individual patients and on the prevalence of mutations in candidate cancer genes is ongoing.

Reportable Outcomes

- A portion of our results on this project was published in the Proceedings of the National Academy of Sciences (PNAS) in September 2011:

  Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, Rieder MJ, Nickerson DA, Corey E, Lange PH, Morrissey C, Vessella RL, Nelson PS, Shendure J. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci U S A. 2011 Oct 11;108(41):17087-92. Epub 2011 Sep 26. PubMed PMID: 21949389; PubMed Central PMCID: PMC3193229.

Conclusions

In summary, by performing exome sequencing of 23 tumors representing a spectrum of aggressive advanced prostate cancers, we identified a large number of previously unrecognized gene coding variants with the potential to influence tumor behavior. However, our results also indicate that with notable exceptions, very few genes are mutated in a substantial fraction of tumors. Furthermore, while the overall mutation frequencies approximate those found in other cancers of epithelial origin, we also identified a distinct subset of tumors that exhibit a hypermutated genome. Ongoing work is directed at performing targeted sequencing of candidate cancer genes in additional samples to establish prevelance of somatic mutations in

each gene as well as patterns of mutational history, as well as at sequencing additional exomes to assess prevalence and identify further lesions involved in advanced stage disease. Furthermore, our results to date illustrate how individual cases can provide information that is potentially therapeutically relevant to a subset of prostate cancer patients. For example, the potentially actionable mutations in BRAF in one patient, as well as the observed heterogeneity with respect to the nature of AR mutation driving therapeutic resistance in different metastases of another patient.

References

Bustamante, C. D., A. Fledel-Alon, et al. (2005). "Natural selection on protein-coding genes in the human genome." Nature **437**(7062): 1153-1157.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics (Oxford, England) **25**(14): 1754-1760.

Li, H., B. Handsaker, et al. (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lohmueller, K. E., A. R. Indap, et al. (2008). "Proportionally more deleterious genetic variation in European than in African populations." Nature **451**(7181): 994-997.

Ng, S. B., E. H. Turner, et al. (2009). "Targeted capture and massively parallel sequencing of 12 human exomes." Nature **461**(7261): 272-276.

O'Roak, B. J., P. Deriziotis, et al. (2011). "Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations." Nat Genet.

Shendure, J. and H. Ji (2008). "Next-generation DNA sequencing." Nat Biotechnol **26**(10): 1135-1145.

Appendices

A published manuscript related to this work is provided as the sole appendix on the pages that follow.

Kumar A, White TA, MacKenzie AP, Clegg N, Lee C, Dumpit RF, Coleman I, Ng SB, Salipante SJ, Rieder MJ, Nickerson DA, Corey E, Lange PH, Morrissey C, Vessella RL, Nelson PS, Shendure J. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. Proc Natl Acad Sci U S A. 2011 Oct 11;108(41):17087-92. Epub 2011 Sep 26. PubMed PMID: 21949389; PubMed Central PMCID: PMC3193229.